

# Binding Social and Cultural Networks: A Model

Camille Roth\*, Paul Bourguine\*

6th February 2008

## Abstract

Until now, most studies carried onto social or semantic networks have considered each of these networks independently. Our goal here is to bring a formal frame for studying both networks empirically as well as to point out stylized facts that would explain their reciprocal influence and the emergence of clusters of agents, which may also be regarded as “*cultural cliques*”. We show how to apply the Galois lattice theory to the modeling of the *coevolution* of social and conceptual networks, and the characterization of cultural communities. Basing our approach on Barabasi-Albert’s models, we however extend the usual preferential attachment probability in order to take into account the reciprocal influence of both networks, therefore introducing the notion of dual distance. In addition to providing a theoretic frame we draw here a program of empirical tests which should give root to a more analytical model and the consequent simulation and validation. In a broader view, adopting and actually implementing the paradigm of cultural epidemiology, once we have understood network formation and evolution, we could therefore proceed further with the study of knowledge diffusion and explain how the social network structure affects concept propagation and in return how concept propagation affects the social network.

## Introduction

Many studies have been carried on real networks, considering them as complex systems and trying to explain their formation and dynamics [1, 12]. Whereas the models proposed have initiated efficient proposals for explaining the general properties of these networks (mostly about node degrees and in particular the broadly shared *scale-free* property [3, 19]), yet they often lack robust explanations for “advanced” topological features such as clustering [13] – a feature especially observed in social networks which denotes the

propensity of two agents to be connected together if they have common acquaintances [20]. Until recently though, no attempt had been made to treat differently social networks in respect of other real networks.

A recent study by Newman & Park [14] however points out social networks singularities regarding the correlations in degrees of adjacent vertices as well as the clustering structure. Trying to model the way beliefs propagate among social networks of agents (in our case, a community of scientists), that is, explain how the social network structure affects concept propagation and in return how concept propagation affects the social network, we will consider here another approach stemming from a social psychology argument: attraction for same-profile people (“homophily”) is indeed key in the formation of social acquaintances [11].

Apart from properties relative to the social network such as node degrees, we can assume that the dominant criterion for choosing a scientific partner mostly depends on the cultural similarity of two agents. An economic model of knowledge creation developed in [6] already tries to take into consideration agents profile (elements of a vector space) in order to explain the structure of the economic network – agents match two by two to produce new knowledge according to their profile.

We introduce here a network dual to the social network, the network of cultural representations, denoted as concepts. Our goal is to bring a formal frame for studying both networks empirically as well as to point out stylized facts that would explain their reciprocal influence and the emergence of clusters of agents, which may also be regarded as “*cultural cliques*”. In a broader view, adopting and actually implementing the paradigm of cultural epidemiology [7, 18], we could therefore proceed further with the study of knowledge diffusion.

\*CREA (Center for Research in Applied Epistemology), CNRS/Ecole Polytechnique, 1 rue Descartes, 75005 Paris, France. Corresponding author: [roth@poly.polytechnique.fr](mailto:roth@poly.polytechnique.fr)

# 1 Networks

We present hereafter the networks we work with – social and conceptual – as well as the links within and between them.

## 1.1 Social network

**Definition 1.** *The social network  $\mathcal{S}$  is represented by the network of coauthorship, where nodes are authors and links are collaborations.*

Thus  $\mathcal{S} = (\mathbf{S}, \lambda_{\mathbf{S}})$ , where  $\mathbf{S}$  denotes the set of authors and  $\lambda_{\mathbf{S}}$  denotes the set of undirected links. As time evolves, new articles are published, new nodes are possibly added to  $\mathbf{S}$  and new links are created between each pair of co-authors. We actually consider the temporal series of networks  $\mathcal{S}(t)$  with  $t \in \mathbb{N}$  (articles are published with a date, thus an integer), for we want to observe the dynamics of the network. Yet we will usually omit the reference to  $t$  because  $\mathcal{S}$  always depends implicitly on time.

An important question of design here is the nature of links. Depending on the model goals and precision, we may want to take into account the fact that two nodes have co-authored more than one paper (thus introducing *link strength*), or that their collaborations are more or less recent (thus introducing *link age*). Indeed, an empirical study of paper citation distribution [16] shows that the probability of citation decreases in respect of time, since papers are gradually forgotten or obsolete; while another model examining the world wide web network [10] notices that the link distribution must depend on the time that has elapsed since a web site was created.

**Weighted networks** Relationships should consequently be different according to whether agents have collaborated only once and a long time ago, or they have recently co-authored many articles. An easy and practical way for dealing with these notions is to use a weighted network:

- in a *non-weighted network*, we say that two nodes are linked as soon as there exists one coauthored article. Links can only be active *or* inactive.
- in a *weighted network*, links are provided with a weight  $w \in \mathbb{R}^+$ , possibly evolving in time. We can therefore easily represent multiple collaborations by increasing the weight of a link, or render the age of a relationship by decreasing this weight (for instance by applying an aging function).

This method enables us to model a non-weighted network by assigning weights of 1 or 0 respectively to active or inactive links. This method also leaves room for creating *ex post* a non-weighted network from a weighted network by setting a threshold, such that a link is active when its weight exceeds the threshold, otherwise inactive.

## 1.2 Conceptual network

The conceptual network is very similar to the social network - and as we will see, dual:

**Definition 2.** *The conceptual network  $\mathcal{C}$  is the network of joint appearances of concepts within articles, where nodes are concepts and links are co-occurrences.*

Identically to  $\mathcal{S}$ , we have  $\mathcal{C} = (\mathbf{C}, \lambda_{\mathbf{C}})$ . When a new article appears, new concepts are possibly added to the network, and new links are added between co-appearing concepts. Here again, as in the case of the social network, one may use a weighted network to render the frequency or the age of co-occurrences.

However, the whole point is now to define precisely what a *concept* is. Is it a paradigm like “*universal gravitation*”, a scientific field like “*molecular biology*”, or a simple word like “*interferon*”? In particular, what is a concept such that we can observe its appearance in an article?

This notion needs be not too precise nor too wide. For instance, authors provide their articles with keywords: apparently, considering these keywords as concepts seems to constitute a relevant level of categorization while being a convenient idea. However, such keywords have not proven to be very reliable indicators of the issues articles are dealing with, for authors often omit important keywords or specify poorly relevant ones.

**Words as concepts** The idea would be to create new keywords from the words appearing in articles, and at first we will say that *each word is a concept*. This definition does not prevent us from observing higher-level concepts such as scientific fields or paradigms, since we can easily refer to these concepts *a posteriori* by considering sets of strongly connected words. For example, we could interpret the set of frequently co-occurring words {“*cell*”, “*cancer*”, “*DNA*”, “*gene*”, “*genetic*”, “*genetics*”, “*molecular*”} as *molecular biology*. This understanding refers to the notion of *meme* introduced by Dawkins [7].

Moreover, we proceed only with words present in what we consider to be the most relevant article data: the title and the abstract. We prefer to set aside article content, since first and above all it is rarely available; second, it could make appear too many very precise though irrelevant words. Therefore, we assume that all important concepts an article tackles and bears on are explicitly used in its title or its abstract. Of course, we also need define a list of words to be ignored, or “*stop words*”, including grammatical and un-significant words (“*is*”, “*with*”, “*study*”, etc.) as well as non-discriminating words (e.g., “*biology*” within a community of biologists) for which a robust criterion will be proposed in §2.3.

### 1.3 Binding the two networks

As the social network is the network of joint appearances of authors, so is the conceptual network with concepts, establishing an obvious duality between the two networks. This duality is key if we want to bind them and explain their reciprocal influence.

In the same way we did with the previous networks, we link scientists to the words they use, i.e. we add a link whenever an author and a word co-appear within an article.

Hence considering the two networks  $S$  and  $C$ , we deal with three kinds of quite similar links: (i) between pairs of scientists, (ii) between pairs of concepts, and (iii) between concepts and scientists; thus setting up three kinds of binary relations:

- (i) a set of symmetrical relations  $\mathcal{R}_\alpha^S \subset S \times S$  from the social network to the social network, and such that given  $\alpha \in \mathbb{R}$  and two scientists  $s$  and  $s'$ , we have  $s \mathcal{R}_\alpha^S s'$  iff the link between  $s$  and  $s'$  has a weight  $w$  strictly greater than the threshold  $\alpha$ .
- (ii) a set of symmetrical relations  $\mathcal{R}_\alpha^C \subset C \times C$  from the conceptual network to the conceptual network, and such that given  $\alpha \in \mathbb{R}$  and two concepts  $c$  and  $c'$ ,  $c \mathcal{R}_\alpha^C c'$  iff the link between  $c$  and  $c'$  has a weight  $w > \alpha$ .
- (iii) a binary relation  $\mathcal{R}_\alpha \subset S \times C$  from the social network to the conceptual network, and such that given  $\alpha \in \mathbb{R}$ , an author  $s$  and concept  $c$ ,  $s \mathcal{R}_\alpha c$  iff the link between  $s$  and  $c$  has a weight  $w > \alpha$ .

Let us examine the special case  $\alpha = 0$ . Noticing that  $\alpha < \alpha' \Rightarrow \mathcal{R}_{\alpha'}^{(\cdot)} \subset \mathcal{R}_\alpha^{(\cdot)}$ , thus giving  $\forall \alpha > 0, \mathcal{R}_\alpha^{(\cdot)} \subset \mathcal{R}_0^{(\cdot)}$ , we infer that the relations  $\mathcal{R}_0^{(\cdot)}$  are maximal, i.e. two nodes are related whenever there

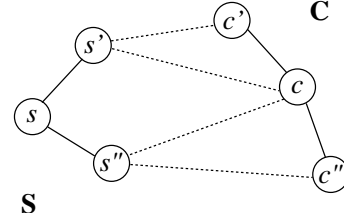


Figure 1: Sample network with  $S = \{s, s', s''\}$ ,  $C = \{c, c', c''\}$ , and relations  $\mathcal{R}^S$ ,  $\mathcal{R}^C$  (solid lines) and  $\mathcal{R}$  (dashed lines).

exists a link binding them, whatever its weight. To ease the notation, we will identify  $\mathcal{R}_0^S$  to  $\mathcal{R}^S$ ,  $\mathcal{R}_0^C$  to  $\mathcal{R}^C$ , and  $\mathcal{R}_0$  to  $\mathcal{R}$ .

## 2 Lattices and epistemic closure

The basic ingredients being defined, we need yet another formal tool to formulate stylized facts about knowledge and people conveying it. Galois lattices appear to be a suitable frame to describe these facts as they offer a powerful structure for concept categorization. They are also being therefore widely used in conceptual knowledge systems [22] and formal concept classification [9]. In the field of social networks, White & Freeman have already explored an application of this theory to social networks [8], though his model deals with agents and social events they attend. The goal of this section is to present the Galois lattice theory and show how we can use it here to describe efficiently the relationships between  $S$  and  $C$ .

### 2.1 Sets and relations

Let us first consider two finite sets  $A$  and  $B$  between which we have a binary relation  $R \subseteq A \times B$ . We introduce the operation “ $\wedge$ ” such that for any element  $x \in A$ ,  $x^\wedge$  is the set of  $B$  elements  $R$ -related to  $x$ . Extending this definition to subsets  $X \subseteq A$ , we denote by  $X^\wedge$  the set of  $B$  elements  $R$ -related to every element of  $X$ , namely:

$$x^\wedge = \{y \in B \mid xRy\} \quad (1a)$$

$$X^\wedge = \{y \in B \mid \forall x \in X, xRy\} \quad (1b)$$

Similarly, “ $\star$ ” is the dual operation so that  $\forall y \in B$ ,

$$\forall Y \subseteq B,$$

$$y^* = \{x \in A \mid xRy\} \quad (2a)$$

$$Y^* = \{x \in A \mid \forall y \in Y, xRy\} \quad (2b)$$

By definition we set  $(\emptyset)^\wedge = B$  and  $(\emptyset)^* = A$ .

These operations enjoy the following properties:

$$X \subseteq X' \Rightarrow X'^\wedge \subseteq X^\wedge \quad (3a)$$

$$Y \subseteq Y' \Rightarrow Y'^* \subseteq Y^* \quad (3b)$$

and

$$X \subseteq X^{\wedge*} \quad (4a)$$

$$Y \subseteq Y^{\star\wedge} \quad (4b)$$

Also, we have<sup>1</sup>:

$$(X \cup X')^\wedge = X^\wedge \cap X'^\wedge \quad (5a)$$

$$(Y \cup Y')^* = Y^* \cap Y'^* \quad (5b)$$

**Closure operation** More important, the following property holds true,<sup>2</sup>

$$((X^\wedge)^*)^\wedge = X^\wedge \text{ and } ((Y^*)^\wedge)^* = Y^* \quad (6)$$

and therefore we are enabled to define the operation “ $\wedge^*$ ” as a *closure operation* [5], in that it is:

$$\text{extensive,} \quad X \subseteq X^{\wedge*} \quad (7a)$$

$$\text{idempotent} \quad (X^{\wedge*})^{\wedge*} = X^{\wedge*} \quad (7b)$$

$$\text{and increasing.} \quad X \subseteq X' \Rightarrow X^{\wedge*} \subseteq X'^{\wedge*} \quad (7c)$$

We say that  $X$  is a *closed* subset if  $X^{\wedge*} = X$ .

## 2.2 Galois lattices

We need now consider the set of couples of subsets of  $A$  and  $B$  and build a new structure onto it.

**Complete couples** Given two subsets  $X \subseteq A$  and  $Y \subseteq B$ , a couple  $(X, Y)$  is said to be *complete* iff  $Y = X^\wedge$  and  $X = Y^*$ .

Yet such a couple is actually a  $(X, X^\wedge)$  where  $X^{\wedge*} = X$ . Therefore, complete couples correspond obviously to couples of subsets of  $A$  and  $B$  closed under  $\wedge^*$ . This will allow us to define a new kind of lattice from  $A$ ,  $B$  and  $R$ . We first recall the definition of a *lattice*:

<sup>1</sup>And accordingly,  $X^\wedge = (\bigcup_{x \in X} \{x\})^\wedge = \bigcap_{x \in X} x^\wedge$ .

<sup>2</sup>Indeed, (3a) applied to (4a) leads to  $(X^{\wedge*})^\wedge \subseteq X^\wedge$ , while (4b) applied to  $X^\wedge$  gives  $(X^\wedge)^\wedge \subseteq (X^\wedge)^{\wedge*}$ .

**Definition 3.** A set  $(L, \sqsubseteq, \sqcup, \sqcap)$  is a lattice if every finite subset  $H \subseteq L$  has a least upper bound in  $L$  noted  $\sqcup H$  and a greatest lower bound in  $L$  noted  $\sqcap H$  under the partial-ordering relation  $\sqsubseteq$ .

In this respect the set of subsets of a set  $X$  provided with the usual inclusion, union and intersection,  $(\mathcal{P}(X), \subseteq, \cup, \cap)$ , is a lattice. Any partially-ordered finite set is also a lattice, and so is a *Galois lattice* [4]:

**Definition 4.** Given a relation  $R$  between two finite sets  $A$  and  $B$ , the Galois lattice  $\mathcal{G}_{A,B,R}$  is the set of every complete couple  $(X, Y) \subseteq A \times B$  under relation  $R$ . Thus,

$$\mathcal{G}_{A,B,R} = \{(X^{\wedge*}, X^\wedge) \mid X \subseteq A\} \quad (8)$$

Indeed  $\mathcal{G}_{A,B,R}$  is finite and is provided with the following natural partial order  $\sqsubseteq$ :

$$(X, X^\wedge) \sqsubseteq (X', X'^\wedge) \Leftrightarrow X \subseteq X' \quad (9)$$

**Formal concept lattice** As Wille points out in [22], this structure constitutes a solid formalization of the philosophical apprehension of a concept characterized by its *extension* (the physical implementation or the group of things denoted by the concept) and its *intension* (the properties or the internal content of the concept).

In a pair  $g = (X, X^\wedge)$  considered as a formal concept,  $X$  may be seen as the extension of  $g$  while  $X^\wedge$  is its intension. For a given  $X \subseteq A$ ,  $X^\wedge$  will represent the set of properties shared by all objects of  $X$ , whereas for a given set of properties  $Y \subseteq B$ ,  $Y^*$  will be the set of objects of  $A$  actually fulfilling them.

Also, using the strict partial order  $\sqsubset$ , we can talk of *formal subconcept* by saying  $g$  is a subconcept of  $g'$  iff  $g \sqsubset g'$ . Hence  $g$  can be seen as a specification of  $g'$ , since the number of its properties increases ( $X^\wedge \supset X'^\wedge$ , thus defining  $g$  more precisely) while less objects belongs to its extension ( $X \subset X'$ ). Conversely,  $g'$  is a “*superconcept*” or a generalization of  $g$ ; we have so a tool of generalization and specification of formal concepts [21].<sup>3</sup>

## 2.3 Applying lattices to $\mathcal{S}$ and $\mathcal{C}$

It is possible now to give some semantics to these tools with respect to our networks  $\mathcal{S}$  and  $\mathcal{C}$ . For this purpose, we will consider the two finite sets  $\mathbf{S}$ ,  $\mathbf{C}$ , the relation  $\mathcal{R}$  and  $\mathcal{G}_{\mathbf{S},\mathbf{C},\mathcal{R}}$ .

<sup>3</sup>Of course, these notions are dually defined, i.e. it is possible to consider  $Y$  as an extension and  $Y^*$  as an intension.

First, for an author  $s \in \mathbf{S}$ ,  $s^\wedge = \{c \mid s\mathcal{R}c\}$  represents the set of the concepts he talked about or the fields he dealt with. Proceeding identically with a concept  $c \in \mathbf{C}$ ,  $c^\star = \{s \mid s\mathcal{R}c\}$  represents the set of scientists who used the concept  $c$  in at least one of their papers.

Then, for a group of authors  $S \subseteq \mathbf{S}$ ,  $S^\wedge$  represents the words being used by every author  $s \in S$ , while for a set of words  $C \subseteq \mathbf{C}$ ,  $C^\star$  is the set of agents using every concept  $c \in C$ . Moreover, we can easily derive from (5) the words used by a community  $S \cup S'$  by taking the intersection  $S^\wedge \cap S'^\wedge$ , or the authors corresponding to the merger of any two sets of concepts  $C \cup C'$  by taking  $C^\star \cap C'^\star$ .

An example is shown on figure 2. For instance,  $s_4^\wedge = \{c_1, c_4, c_5\}$  and  $\{c_1, c_6\}^\star = \{s_3, s_5\}$ . If we consider the matrix  $R$  representing relation  $\mathcal{R}$  as follows,

$$R = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

where  $R_{i,j}$  is non-zero when  $s_i \mathcal{R} c_j$ , we can easily read  $s_i^\wedge$  on rows and  $c_j^\star$  on columns.

**Epistemic closure and epistemic categories** Seeing concepts as *properties* of authors who use them (skills in scientific fields as cognitive properties) and authors as *extensions* of concepts (implementation of concepts within authors), one can make a very fertile usage of the lattice  $\mathcal{G}_{\mathbf{S}, \mathbf{C}, \mathcal{R}}$  by setting up an epistemic taxonomy with the help of formal concepts made of couples  $(S, C)$  with  $S \subseteq \mathbf{S}$ ,  $C \subseteq \mathbf{C}$ . We may indeed consider such formal concepts as *schools of thought* constituted by the community of agents  $S$  working and writing on the field  $C$ , a formal subconcept simply being a trend inside a school. By community we understand henceforth *epistemic community*, that is to say neither a department nor a group of research.

In addition, we recall that for such a complete couple from the Galois lattice,  $C = S^\wedge$ ,  $S = C^\star$  and finally  $S = S^{\wedge\star}$ . What does  $S^{\wedge\star}$  actually represent? It is the set of scientists using *at least* the same words as  $S$ . But “ $\wedge\star$ ” being a closure operation,  $S^{\wedge\star}$  closes the set  $S$  by returning all the scientists related to every concept shared among  $S$  – once and for all from (7b) – which makes us call it an *epistemic closure operation*.<sup>4</sup>

<sup>4</sup>Note that given  $S^\wedge = \{c_1, \dots, c_n, c\}$  and  $S'^\wedge = \{c_1, \dots, c_n, c'\}$ , we have  $S' \not\subseteq S^{\wedge\star}$ ,  $S'$  not being in the epistemic closure of  $S$ , which might look quite strange for a human eye who

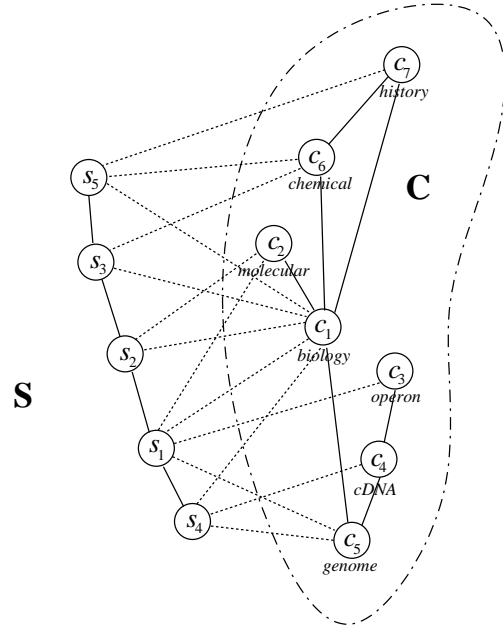


Figure 2:  $\mathbf{S} = \{s_1, s_2, s_3, s_4, s_5\}$ ,  $\mathbf{C} = \{c_1, c_2, c_3, c_4, c_5, c_6, c_7\} = \{\text{biology, molecular, operon, cDNA, genome, chemical, history}\}$  – solid lines:  $\mathcal{R}^{\mathbf{S}}$  and  $\mathcal{R}^{\mathbf{C}}$ , dashed lines:  $\mathcal{R}$ .

Admittedly, for a single scientist  $s$ ,  $s^{\wedge\star}$  will certainly be equal to  $s$ , since there are strong chances that  $\forall s' \in \mathbf{S}, \exists w \in s^\wedge$  and  $w \notin s'^\wedge$ . Considering however a subset  $S \subseteq \mathbf{S}$ , as its cardinal increases there are more and more chances that the closure of  $S$  reaches an actual community of researchers. We conjecture that there is a relevant level of closure for which a set  $S^{\wedge\star}$ , and identically  $C^{\wedge\star}$ , is representative of a field or a trend. This idea is to be compared to Rosch’s basic-level of categorization [17].

$\mathcal{G}_{\mathbf{S}, \mathbf{C}, \mathcal{R}}$  contains all complete couples: this includes naturally most singletons ( $s^{\wedge\star}$ ,  $s^\wedge$ ) as well as  $(\mathbf{S}, \mathbf{S}^\wedge)$ , but also and especially all the intermediary pairs of closed sets. For this purpose, there must then be a gap between couples whose  $S$  is of very small size and those with medium-sized  $S$ , with very few complete couples inbetween; and likewise a gap between

would have said their domains of interest similar.

Another property may help understand better what this closure actually tallies with: given  $S^\wedge = \{c_1, \dots, c_n\}$  and  $S'^\wedge = \{c'_1, \dots, c'_n\}$  such that  $\forall (i, j) \in \{1, \dots, n\}^2, c_i \neq c'_j$ , we have  $(S \cup S')^{\wedge\star} = \mathbf{S}$ : the closure of two sets of scientists working on totally different issues is the whole community  $\mathbf{S}$  – “there is no way to distinguish  $S$  and  $S'$  from each other with respect to the rest of the community”.

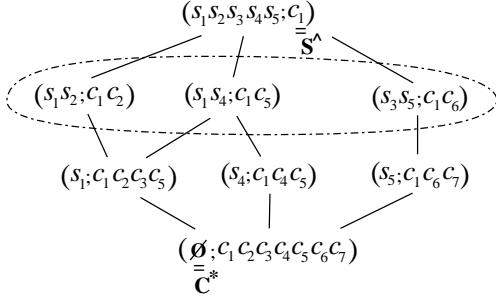


Figure 3: Representation of the whole Galois lattice of our example – the hierarchy is drawn according to the partial order  $\sqsubset$ , i.e. “bottom” $\sqsubset$ “top”. The cultural background  $S^\wedge$  is reduced to “biology”. On the medium-level, we find formal concepts  $(s_1, s_2 ; \text{“biology”}, \text{“molecular”})$ ,  $(s_1, s_4 ; \text{“biology”}, \text{“genome”})$ ,  $(s_3, s_5 ; \text{“biology”}, \text{“chemical”})$ .

couples with small- and medium-sized  $C$  sets. This medium level shall constitute our basic-level of epistemic categorization, whereas above it (“superordinate categories”) the field would be too general, and too precise under it (“subordinate categories”).

If we define the *epistemic family* of an agent as the set of (possibly many) complete couple(s) which he is a member of, and also whose  $S$  set size is above a certain threshold, it could be very useful to identify the basic-level epistemic categories to help fit the threshold value.

**Cultural background** Interestingly,  $S^\wedge$  represents the concepts the whole community shares – the “background” – and are obviously too common to be discriminating. This set could actually constitute a very appropriate companion to the list of stop words we mentioned in §1.2. On the other hand,  $C^*$  does not enjoy in general any such property and is empty; the contrary would mean that there would be at least one author having used *every* word in use among the whole community, which would be quite dreadful in fact.

### 3 Applications

This section is devoted to pointing out the joint applications of the two preceding sections to the observation, description and eventually model of the dynamics of our networks.

#### 3.1 Network dynamics

We will first try to account for the network evolution by extending already existing models to make them include the improvements offered by the theoretic frame exposed above. The growing-network model proposed by Barabasi & Albert [2] will be our basis. It is directed by two key phenomenas: (i) a *constant rate of growth* (the number of nodes at any time  $t$  is  $\alpha t$ ), justified by the fact that real networks “grow by the continuous addition of new nodes” [1]; and (ii) a *preferential attachment* – external (new nodes join the system) as well as internal (links appearing between existing nodes) – however neglecting aging considerations. This is borne out by the preference one may for example exhibit towards an already well-connected agent, as being more recognized, famous, reliable or simply efficient.

If we assume yet that homophily is essential to the system dynamics, the preferential attachment must be modified in order to take into account similarity between agents or between concepts: nodes will indeed join preferentially more connected but also more similar nodes. Thus, the preferential attachment probability of a node to another node within the same network, usually denoted by  $\Pi(k_1, k_2)$  where  $k_1, k_2$  are the degrees of the nodes connecting to each other<sup>5</sup>, should not be uniform with respect to the nodes proximity within their dual network. That is to say for instance that  $\Pi$  should depend for a scientist  $s \in S$  both (i) on the degree of other scientists  $s' \in S$  (using  $\mathcal{R}^S$ ) and (ii) on the “distance” between  $s^\wedge$  and  $s'^\wedge$  – or *dual distance* between  $s$  and  $s'$  (using  $\mathcal{R}^C$  and  $\mathcal{R}$ ).

**Dual distance** The notion of dual distance  $d$  needs nonetheless be defined more precisely: in order to measure the similarity or equivalently the difference between two nodes dual sets we will adopt a formula inspired from the notion of Hamming distance on sets of bits.

**Definition 5.** Given  $(s, s') \in S^2$  and using the symmetric difference of two sets we define the dual dis-

<sup>5</sup>This concerns internal attachment. External attachment is, of course, undefined as regards the dual distance.

tance  $d(s, s') \in [0; 1]$  such that:<sup>6</sup>

$$d(s, s') = \frac{|(s^\wedge \setminus s'^\wedge) \cup (s'^\wedge \setminus s^\wedge)|}{|s^\wedge \cup s'^\wedge|}$$

We expect to find a different behavior with respect to this parameter  $d$ : indeed if cliques do exist, preferential attachment should also depend on a parameter precisely related to the “*cliquishness*”. Newman in [13] considers the number of common acquaintances as an explanatory argument for clique formation. Instead, we will assume that collaborations do essentially occur on account of homophily, while this assumption does not contradict Newman’s argument: two agents are all the more likely to have the same profile that they share many acquaintances.<sup>7</sup>

The probability  $\Pi(k_1, k_2)$  encloses this information without enabling us to discriminate the effect of the dual proximity. We next consider  $\Pi(k_1, k_2, d)$  which takes into account this second variable  $d$ . The main direction to be explored would be to build  $\Pi(k, d)$  on the dual distance defined above. Assuming the independence of  $(k_1, k_2)$  and  $d$ ,  $\int_0^1 \Pi(k_1, k_2, \delta) \rho(\delta) d\delta = \Pi(k_1, k_2)$  holds true, where  $\rho$  is the density of  $d$ .<sup>8</sup>

In any case, a first step will be to determine empirically the shape of  $\Pi$  so that we can infer fertile intuitions for designing an analytical value for  $\Pi$ . Also, although not detailed here, the reasoning holds the same for the preferential attachment in **C**.

### 3.2 *Cliquishness* and coalitions

As regards the cliquishness in particular, another point of interest is to see whether network cliques correspond to closed sets, i.e. whether a  $\mathcal{S}$ -clique is also a  $\wedge\star$ -clique, and whether a  $\mathcal{C}$ -clique is also a  $\star\wedge$ -clique.

<sup>6</sup>Written in a more explicit manner, given two sets  $s^\wedge = \{c_1, \dots, c_n, c_{n+1}, \dots, c_{n+p}\}$  and  $s'^\wedge = \{c_1, \dots, c_n, c'_{n+1}, \dots, c'_{n+q}\}$ ,  $d(s, s') = \frac{p+q}{p+q+n}$ ;  $n$  and  $p, q$  representing respectively the number of elements  $s^\wedge$  and  $s'^\wedge$  have in common and have in proper. We also verify that if  $n = 0$  (disjoint sets),  $d = 1$ ; if  $n \neq 0, p = q = 0$  (same sets),  $d = 0$ ; and if  $s^\wedge \subset s'^\wedge$  (included sets),  $d = \frac{q}{q+n}$ . It is besides easy though cumbersome to show that  $d(\cdot, \cdot)$  is actually a distance.

<sup>7</sup>However another yet better model would also take into account this property and would express it through enhancing our definition of  $d$ .

<sup>8</sup>We could also draw out the distribution  $\Pi(k_1, k_2, d)$  depending on whether agents are member of the same epistemic family (introduced in §2.3) or not. The variable  $d$  would thus belong to  $\{0; 1\}$  and we would actually deal with only two distributions. This option is less robust than the previous one for it relies on a quite “fuzzy” parameter (epistemic family membership), whereas it could offer a more schematic and stylized interpretation.

In other words, we want to know whether schools of thought and scientific fields are also socially and conceptually strongly linked or not. Though we could expect this to be true in real world networks, it is certainly not a fortuitous property for it relies on two different kinds of tools (epistemic closure vs. single network connectivity). We might also want to adopt here an extended definition of a clique (as a fully connected triplet of nodes), and for instance use  $k$ -connectivity<sup>9</sup> [15] (the smallest number of nodes one needs to withdraw from a connected (sub)graph to get a disconnected one).

### 3.3 Sensibility to parameter $\alpha$

At the beginning of our paper (§1.3) we left room for a parameter  $\alpha$  in relations  $\mathcal{R}_\alpha^{(\cdot)}$ , which had been until now implicitly set to 0. Its main function is actually to prevent insignificant links (too old or too rare) to be taken into account: indeed, under a certain threshold of strength or significance, a link would be excluded from  $\mathcal{R}_\alpha^{(\cdot)}$ .

In the extreme case, for  $\alpha$  big enough there is no connection at all. As for the appearance of a giant component in random networks [12], there may be a transition value  $\alpha_c$  above which almost no connectivity exists and under which the network is significantly connected. This hypothesis ought also to be checked.

## Conclusion

Until now, most studies carried onto social networks or conceptual (semantic) networks have considered each of these networks independently. We proposed here a frame for binding them and pointing out their very duality as well as expressing stylized facts about them. The Galois lattice theory has proved useful in helping introduce key notions such as epistemic closure and basic-level of categorization of a scientific field, and in general for characterizing scientific communities.

Next we showed how to apply this structure to the model of the *coevolution* of the social and cultural networks. We have mostly based our approach on Barabasi-Albert’s models, stressing out constant growth and preferential attachment. However instead of considering social and conceptual networks separately, we modified the usual preferential attachment probability in order to take into account the reciprocal

<sup>9</sup>We acknowledge fruitful remarks from Douglas R. White on this point.

influence of both networks. We therefore introduced the notion of dual distance.

Finally, more than providing a theoretic frame, we have in fact also drawn a program of empirical tests which should give root to a more analytical model and the consequent simulation and validation. This step constitutes the first milestone of a broader attempt to implement the paradigm of cultural epidemiology: we could thus describe and explain propagation of concepts through the social network *as well as* propagation of scientists through the conceptual network.

---

## References

- [1] R. Albert & A.-L. Barabasi (2002), *Statistical Mechanics of Complex Networks*, Reviews of Modern Physics, 74:47–97, 01/2002.
- [2] A.-L. Barabasi, H. Jeong, R. Ravasz, Z. Neda, T. Vicsek & A. Schubert (2002), *Evolution of the social network of scientific collaborations*, Physica A 311–590.
- [3] A.-L. Barabasi, Z. Dezso, Z. Oltvai, E. Ravasz and S.-H. Yook (2003), *Scale-free and Hierarchical Structures in Complex Networks*, to appear in Sitges Proceedings on Complex Networks, 2004.
- [4] M. Barbut & B. Monjardet (1970), *Ordre et Classification*, Algèbre et Combinatoire Tome II, Hachette, Paris.
- [5] G. Birkhoff (1948), *Lattice Theory*, American Mathematical Society, Providence, RI.
- [6] R. Cowan, N. Jonard & J.-B. Zimmermann (2002), *The Joint Dynamics of Networks and Knowledge*, in “Heterogenous Agents, Interactions and Economic Performance”, Cowan & Jonard (Eds.), Springer, Lecture Notes in Economics and Mathematical Systems, 521, 155–174.
- [7] R. Dawkins (1976), *Memes, The New Replicator*, in The Selfish Gene, chap. 11, Oxford University Press.
- [8] L. C. Freeman & D. R. White (1993), *Using Galois Lattices to Represent Network Data*, Sociological Methodology, 23:127–146.
- [9] R. Godin, H. Mili, G. W. Mineau, R. Missaoui, A. Arfi & T.-T. Chau (1998), *Design of class hierarchies based on concept (Galois) lattices*, Theory and Practice of Object Systems (TAPOS), 4(2):117–134.
- [10] B. Huberman & L. Adamic (1999), *Evolutionary dynamics of the world wide web*, Nature, September 1999.
- [11] M. McPherson & L. Smith-Lovin (2001), *Birds of a Feather: Homophily in Social Networks*, Annual Review of Sociology, 27:415–440.
- [12] M. E. J. Newman, S. H. Strogatz & D.J. Watts (2000), *Random Graphs with Arbitrary Degree Distributions and their Applications*, arXiv:cond-mat/0007235.
- [13] M. E. J. Newman (2001), *Clustering and Preferential Attachment in Growing Networks*, arXiv:cond-mat/0104209.
- [14] M. E. J. Newman & J. Park (2003), *Why social networks are different from other types of networks*, arXiv:cond-mat/0305612.
- [15] W. W. Powell, D. R. White, K. W. Koput & J. Owen-Smith (2002), *Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences*, SFI Working Paper.
- [16] S. Redner (1998), *How Popular is Your Paper? An Empirical Study of the Citation Distribution*, European Phys. Journal B 4, 131–134.
- [17] E. Rosch & B. Lloyd (1978), *Cognition and Categorization*, American Psychologist, 44(12):1468–1481.
- [18] D. Sperber (1996), *La contagion des idées*, Odile Jacob, Paris (english text: *Explaining Culture: A Naturalistic Approach*, Blackwell Publishers).
- [19] A. Vazquez (2000), *Knowing a Network by Walking on It: Emergence of Scaling*, arXiv:cond-mat/0006132.
- [20] D. J. Watts & S. H. Strogatz (1998), *Collective Dynamics of “small-world” networks*, Nature 393:440–442.
- [21] R. Wille (1992), *Concept lattices and conceptual knowledge systems*, Computers Mathematics and Applications, 23:493.
- [22] R. Wille (1997), *Conceptual graphs and formal concept analysis*, proceedings of the fourth International Conference on Conceptual Structures, #1257 in Lecture Notes on Computer Science, pp. 290–303, Springer-Verlag, Berlin. <http://citeseer.nj.nec.com/wille97conceptual.html>